DOCUMENT RESUME

ED 060 056 TM 001 142

AUTHOR Stallings, William M.; Anderson, Frances E.
TITLE A Note on Some Characteristics and Correlates of the Meier Art Test of Aesthetic Perception.
INSTITUTION Illinois Univ., Urbana. Office of Instructional Resources.
REPORT NO RR-281
PUB DATE [68]
NOTE 11p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Art Appreciation; Biographical Inventories; *College Students; Correlation; Cultural Environment; Data Analysis; Error Patterns; *Grade Prediction; Grades (Scholastic); Item Analysis; Post Testing; Predictive Validity; Pretests; Rating Scales; Scoring Formulas; Testing Problems; *Test Reliability; *Test Validity
IDENTIFIERS MATAP; *Meier Art Test of Aesthetic Perception

ABSTRACT

The reliability and the predictive and concurrent validity of the MATAP were investigated with the implicit goal of improving the prediction of course grades in the College of Fine and Applied Arts. It was found that reliability and validity coefficients were low, and it was suggested that the scoring system was a source of error variance. (MS)

ED 060056

research report # 281

research report

A Note on Some Characteristics and Correlates of the Meier Art Test of Aesthetic Perception

by

William M. Stallings

and

Frances E. Anderson

measurement and research division
office of instructional resources
university of illinois
urbana, illinois

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Research Report # 281

A Note on Some Characteristics and Correlates of the Meier Art Test of Aesthetic Perception

William M. Stallings

University of Illinois

and

Frances E. Anderson

University of Tennessee

Although the Meier Art Test of Aesthetic Judgement, first published in 1929, has been a subject of extensive research, the Meier Art Test of Aesthetic Perception (MATAP) has been little studied since its publication in 1963. Both Meier Art Tests were intended to measure aspects of aesthetic sensitivity.[1] A third member, Creative Imagination, of the Meier Art Test battery was planned, but its development apparently was halted with Meier's death.

It was the purpose of this study to investigate (a) the reliability of the MATAP and (b) the predictive and concurrent validity of the MATAP--including other measures of artistic ability, course grades, and certain biographical data. One explicit goal of this study was the improvement of predicting course grades in the College of Fine and Applied Arts at the University of Illinois. Architecture, art, landscape architecture, music, theatre, and urban planning are the undergraduate curricula offered by the College of Fine and Applied Arts.

PROCEDURE

Two different groups of subjects were used in this study. One sample consisted of 54 undergraduate students at Indiana University. These students were administered the MATAP and the Child Test of Esthetic Sensitivity (Child, 1962) at the beginning and at the end of an introductory course in art education. Because the MATAP is discussed in a separate section of this paper, only the Child Test of Esthetic

---

1. I.L. Child has provided a convenient definition of aesthetic sensitivity:
   It [esthetic sensitivity] refers to the extent to which a person gives evidence of responding to relevant stimuli in some consistent and appropriate relation to the external standard. Esthetic judgement and esthetic preference [similar to Meier's aesthetic perception] may be viewed as special cases of esthetic sensitivity [Child, 1964, p. 49.]

Sensitivity (CTES) will be described at this point. As used in this study the CTES consisted of 90 pairs of slides of art objects. The subject was required to indicate which one of the pair he preferred. A subject's total score was determined by how his response (preferences) compared with the "keyed" preferences. These "keyed" preferences were simply the consensus of judgments by art experts. The CTES represents a longer form of an instrument developed by Bulley (Bulley, 1951).

The second sample was a group of 169 incoming freshman in the College of Fine and Applied Arts (music majors were excluded) at the University of Illinois. These students took the MATAP, and other measures, as part of the freshman testing program. The other measures included the unpublished Illinois Art Ability Test (See Cronback, 1960, p. 316), biographical data relating to art training, and the American College Testing Program (ACT) battery. At the end of the fall semester certain course grades were collected and an overall grade point average (GPA) was computed for each student.

The publisher's catalogue (Bureau of Educational Research and Service, 1966) gives the following description of the Meier Art Test II (Aesthetic Perception):

> The Meier Art Aesthetic Perception Test is designed to measure individual differences in perception of aesthetic merit of different ways of constructing an art object. This is accomplished by observing four versions of the same work of art, with the subject being required to rank them in order of aesthetic merit. There are fifty plates of test items. The score is the number of placements that agree with the scoring key. The test, essentially a test of observational acuity of aesthetic form, should also have value for testing individuals in scientific research, medicine, and other areas where this capacity is important. [p. 21].

In the Preliminary Manual (1963), Meier elaborated on the rational and procedure used in developing his test of aesthetic perception. Concerning the rationale Meier (1963) wrote:

> Theoretically, the Aesthetic Perception Test has its premise in the greater ability or capacity, as observed in artists and confirmed by them, to observe phenomena (people, behavior, objects, etc.) with considerably greater adequacy than will be experienced by the non-art person [p. 1].

The items themselves were chosen so as to provide "a sampling of world art from ancient to contemporary [Meier, 1963, p. 2]." The key "represents a combination of judgments of a limited number of artists, about 350 art students, teachers of art, and an extensive statistical analysis of the results of the testing on both the experimental form and the published form [Meier, 1963, p. 2]." As mentioned previously, a subject's score is determined by the way in which his rank orderings of the four versions of the fifty items agree with those of the key. One point is given for each correct rank order. Hence a maximum of four points per item--and 200 points for the test--is possible. Although the maximum total score is 200, the raw score to percentile rank conversion table (as given in the Preliminary Manual), has an upper limit of only 112.

Unfortunately, no data on reliability are provided in the Preliminary Manual. Moreover, the only indication of any type of validity is a summary of mean scores for (a) artists (about 90), (b) art students in college and "younger artists groups" (77-85), and (c) high school students enrolled in art courses (72-76). Even though in the Preliminary Manual Meier promised to publish a revised manual and a final key, this apparently had not been done by the fall of 1967.

EVIDENCE FOR RELIABILITY

For the Indiana University sample the correlation between pretest and post-test scores on the MATAP was .220. This low correlation might be explained in part, by the intervening treatment of instruction in art education. However, this is not a satisfactory hypothesis for the following reason. With the same sample over the same time period and intervening treatment the correlation

between pretest and posttest scores of the CTES was .702. Hence it appears that the test-retest reliability of the MATAP is actually quite low.[2]

When data from the University of Illinois sample were item analyzed, each of the four versions (alternatives) of each of the fifty items was treated as a separate true-false item. That is, the subject either selected the keyed rank order or he didn't. Because of the scoring system, the test could be considered a 200 item test. By making this assumption, a Kuder-Richardson formula 20 (KR-20) of .626 and Kuder-Richardson formula 21 (KR-21) of .584 were computed. Clearly, these estimated reliabilities are inflated by the artifically expanded length of the test (200 items in contrast to the original 50). Indeed, if one were to use the Spearman-Brown Prophecy formula to correct the reliability back to a test of 50 items, the KR-20 shrinks to .295 and the KR-21 to .260. Admittedly, this procedure is not psychometically "cricket." Nevertheless, it is obvious that the internal consistency reliability is not high. To get a somewhat different estimate of the internal consistency reliability the variance of total scores was computed. Fifty individual item variances were generated and then summed. By using these data, a "coefficient alpha" (Nunnally, 1968) of .401 was derived.

---

2. Concerning test-retest reliability, Nunnally (1968) has noted: It is recommended that the retest method generally not be used to estimate reliability, but there are some exceptions. In some types of measures, the retest probably would not be markedly affected by the first testing. This would be the case, for example, if an individual were required to rate the pleasantness of 200 designs. The sheer number of ratings would make remembering the ratings of individual designs very difficult, and consequently the retest would be largely independent of the earlier testing. Also, the scores would be more nearly independent if there were a long time between testings....[p. 216].

It should be mentioned that some of the unreliability in the MATAP may be caused by the scoring system. Consider two hypothetical examples: One, the keyed correct rank order (from best to worst) was ABCD, the subject's answer was BADC, and his item score was zero. However, the subject was able to distinguish between the two best and two worst items. Two, the keyed correct rank order (from best to worst) again was ABCD, the subject's answer was BCDA, and his item score was zero. This time the subject had three alternatives in relative correct rank order, but not in absolute correct rank order. Thus a "correct" discrimination and ordering of three of the alternatives was nullified by an incorrect rank ordering of the fourth alternative. That the present scoring scheme is unsatisfactory is implied by the publisher's statement: "Work is in progress toward development of an improved scoring system [Bureau of Educational Research and Service, 1966, p. 21]."

EVIDENCE FOR VALIDITY

As was mentioned in the "procedure" section of this paper a sample of Indiana University art education was given both the MATAP and the CTES on a pretest-posttest basis. For the pretests the correlation between MATAP and the CTES was -.101. The correlation between posttest scores of the MATAP and the CTES was .058. Clearly, these are extremely low correlations for two tests which supposedly measure similar aspects of aesthetic sensibility. One might speculate that the small size of these correlations was caused by the lack of reliability in the MATAP. However, one may substitute test-retest reliabilities (.220 for the MATAP and .702 for the CTES) in the usual formula for correlation corrected for attenuation. The corrected correlation between pretest score of the MATAP and the CTES was -.257, the corrected correlation between posttest score of the same tests was .148. This is not impressive evidence for concurrent validity.

Further evidence concerning the concurrent and predictive validity of the MATAP was obtained from the University of Illinois sample. MATAP scores correlated .279 with scores from the Illinois Test of Art Abilities. The Illinois Art Ability Test (see Cronbach, 1960, p. 316) is a work sample or job-replica type of instrument which has had moderate predictive validity for course grades in art and architecture. With three scores from the American College Testing Program (ACT) battery, the MATAP had the following correlations: (a) .124 with ACT English: (b) -.089 with ACT Mathematics; (c) -.026 with ACT Composite. Even after allowing for the test's apparently low reliability, it would seem that the MATAP is not measuring general scholastic aptitude to any great extent. Nor does the MATAP do well in predicting scholastic achievement, i.e., course grades in art and architecture. Table 1 presents predictive validities of the MATAP with various course grades and with first semester grade-point average (GPA). Incidentally, at the University of Illinois grades in studio art courses are determined largely from the rating of art objects by a faculty jury.

The correlations of MATAP total scores and certain biographical data were relatively high--in comparison with the course grade and MATAP correlations. MATAP scores correlated .162 with a dichotomously scored item on training in art or related work. Moreover, MATAP total scores correlated .218 with the number of years training in art or related work. Length of interest in artistic work and MATAP scores correlated .302. The last two correlations were significantly greater than zero at the .05 level (two-tailed test).

DISCUSSION

The low reliability, both test-retest and internal consistency, of the MATAP is a limiting factor on the size of the validity coefficients. It is possible that the scoring system itself is a source of error variance. Hence an improved

scoring scheme is needed. With the present scoring scheme the limited evidence of validity includes: (a) modest significant correlations with years of art training and years of interest in art (b) different mean scores for artists and non-artists as reported in the manual [Meier, 1963, p. 2], (c) a test format and associated test items that meet Child's definition[3] for a measure of aesthetic preference (perception), (d) the significant, if small, correlation with the Illinois Art Ability Test and the negligible correlations with ACT scores, which indicate, perhaps, that some separate and specific ability was being tapped. The MATAP does not appear to be a promising instrument for improving the prediction of course grades in art related subjects.

---

3. Esthetic preference, as a measured variable, is the extent to which, when a person expresses (by word or action) his relative liking or disliking of various stimuli corresponds to their esthetic value as defined by the external standard [Child, 1964, p. 49].

TABLE 1

VALIDITY COEFFICIENTS[1] OF MATAP SCORES WITH

COURSE GRADES AND FIRST SEMESTER GPA

Variable	Product-Moment Correlation with MATAP
1. MATAP (N = 127)	
2. Engineering Drawing (N = 72)	.012
3. Architectural Design (N = 38)	.074
4. Freehand Drawing (N = 38)	.164
5. Analytic Geometry (N = 74)	.057
6. College Algebra (N = 11)	-.193
7. Drawing (N = 48)	.138
8. Design (N = 48)	-.056
9. Drawing Theory (N = 48)	.022
10. Overall First Semester GPA (N = 126)	.145

1. None of the validity coefficients is significantly greater than zero ($p > .05$, two-tailed test.)

REFERENCES

Bulley, M. H. Art and everyman Vol. 1, London: Batsford, 1951.

Bureau of Educational Research and Service. Standard tests. (NO. 51) The University of Iowa, Iowa City, Iowa: BERS, 1966.

Child, I. L. A Study of esthetic judgment (Cooperative Research Project NO. 669) New Haven: Yale University, 1962.

Child, I. L. Some observations on the meaning of some measures of esthetic sensitivity. Journal of Psychology, 1964, 1957, 1949-1964.

Cronbach, L. J. Essentials of psychological testing, (2nd ed.) New York: Harper, 1960.

Meier, N. C. The Meier Art Tests II, Aesthetic Perception: preliminary manual 1963. Iowa City, Iowa: Bureau of Educational Research and Service, University of Iowa, 1963.

Nunnally, J. Psychometric theory. New York: McGraw-Hill, 1968.